Regression Models Course Project

Nick Pistiolis

Executive Summary

This report explores the relationship between a set of variables and miles per gallon (MPG) (outcome). It also answers the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Data Processing

Load the data.

```
data("mtcars")
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Convert the variable am from numeric to factor "automatic" - "manual".

```
## [1] "factor"

## [1] "automatic" "manual"
```

Exploratory Analysis

Create a boxplot to examine the relationship between mpg and transmission type.

```
plot(mtcars$am,mtcars$mpg,ylab="Miles Per
Galon",xlab="Transmission",main="MPG by Transmission Type")
```

Appendix Graph 1

The boxplot shows better mpg for manual transmission cars, but it examines only the relation between transmission and mpg, assuming all other variables are the same, something that it is not true. So a multivariable regression analysis must be done to compare mpg with other variables that influence mpg.

Multivariable Regression Model

Step function is used to determine which variables (predictors) have the most significance to the outcome (mpg).

```
model<-lm(mpg~.,data = mtcars)
summary(step(model,trace=0))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ammanual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

This shows that in adddition to transmission, weight of the vehicle as well as accelaration speed have the highest relation to explaining the variation in mpg. The adjusted R^2 is 84% which means that the model explains 84% of the variation in mpg indicating it is a robust and highly predictive model. So from the value of R^2, there is no need for interaction terms in the model. Hence it is expected two parallel planes in a 3d space with manual transmission plane, on average, have 2.94 more MPGs.

```
library(car)
library(rgl)
scatter3d(mpg~wt+qsec|am ,data=mtcars)
```

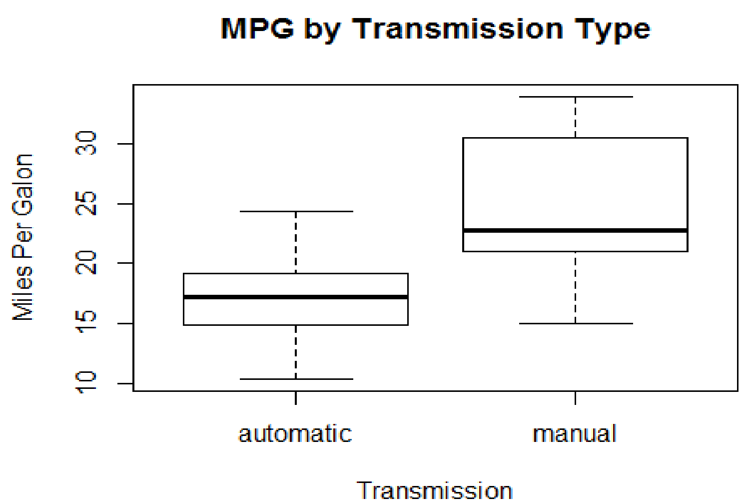
Appendix Graph 2

Results

This model explains 84% of the variance in miles per gallon (mpg). Manual transmission cars have 2.94 MPGs more than automatic transmission cars. However this effect is much lower than when we did not adjust for weight and qsec.

Appendix

Graph 1



Graph 2